NOVEL VECTORS AND GENES EXHIBITING INCREASED EXPRESSION

INSAI

Related Applications

This application claims priority to U.S. Serial No. 60/071,596, filed on January 16, 1998, and to U.S. Serial No.60/067,614, filed on December 5, 1997, the entire contents both of which are incorporated herein by reference.

Background of the Invention

Recombinant DNA technology is currently the most valuable tool known for producing highly pure therapeutic proteins both *in vitro* and *in vivo* to treat clinical diseases. Accordingly, a vast number of genes encoding therapeutic proteins have been identified and cloned to date, providing valuable sources of protein. The value of these genes is, however, often limited by low expression levels.

This problem has traditionally been addressed using regulatory elements, such as optimal promoters and enhancers, which increase transcription/expression levels of genes. Additional techniques, particularly those which do not rely on foreign sequences (e.g., viral or other foreign regulatory elements) for increasing transcription efficiency of cloned genes, resulting in higher expression, would be of great value.

Accordingly, the present invention provides novel methods for increasing gene expression, and novel genes which exhibit such increased expression.

Gene expression begins with the process of transcription. Factors present in the cell nucleus bind to and transcribe DNA into RNA. This RNA (known as pre-mRNA) is then processed via splicing to remove non-coding regions, referred to as introns, prior to being exported out of the cell nucleus into the cytoplasm (where they are translated into protein). Thus, once spliced, pre-mRNA becomes mRNA which is free of introns and contains only coding sequences (i.e., exons) within its translated region.

Splicing of vertebrate pre-mRNAs occurs via a two step process involving splice site selection and subsequent excision of introns. Splice site selection is governed by definition of exons (Berget et al. (1995) *J. Biol. Chem.* 270(6):2411-2414), and begins with recognition by splicing factors, such as small nuclear ribonucleoproteins (snRNPs), of consensus sequences located at the 3' end of an intron (Green et al. (1986) *Annu. Rev. Genet.* 20:671-708). These sequences include a 3' splice acceptor site, and associated branch and pyrimidine sequences located closely upstream of 3' splice acceptor site (Langford et al. (1983) *Cell* 33:519-527). Once bound to the 3' splice acceptor site, splicing factors search downstream through the neighboring exon for a 5' splice donor site. For internal introns, if a 5' splice donor site is found within about 50 to 300 nucleotides downstream of the 3' splice acceptor site, then the 5' splice donor site will generally be selected to define the exon (Robberson et al. (1990) *Mol. Cell. Biol.* 10(1):84-94), beginning the process of spliceosome assembly.

Accordingly, splicing factors which bind to 3' splice acceptor and 5' splice donor sites communicate across exons to define these exons as the original units of spliceosome assembly, preceding excision of introns. Typically, stable exon complexes will only form and internal introns thereafter be defined if the exon is flanked by both a 3' splice acceptor site and 5' splice donor site, positioned in the correct orientation and within 50 to 300 nucleotides of one another.

It has also been shown that the searching mechanism defining exons is not a strict 5' to 3' (i.e., downstream) scan, but instead operates to find the "best fit" to consensus sequence (Robberson et al., *supra.* at page 92). For example, if a near-consensus 5' splice donor site is located between about 50 to 300 nucleotides downstream of a 3' splice acceptor site, it may still be selected to define an exon, even if it is not consensus. This may explain the variety of different splicing patterns (referred to as "alternative splicing") which is observed for many genes.

Summary of the Invention

The present invention provides novel DNAs which exhibit increased expression of a protein of interest. The novel DNAs also can be characterized by increased levels of cytoplasmic mRNA accumulation following transcription within a cell, and by novel splicing patterns. The present invention also provides expression vectors which provide high tissue-specific expression of DNAs, and compositions for delivering such vectors to cells. The invention further provides methods of increasing gene expression and/or modifying the transcription pattern of a gene. The invention still further provides methods of producing a protein by recombinant expression of a novel DNA of the invention.

In one embodiment, a novel DNA of the invention comprises an isolated DNA (e.g., gene clone or cDNA) containing one or more consensus or near consensus splice sites (3' splice acceptor or 5' splice donor) which have been corrected. Such consensus or near consensus splice sites can be corrected by, for example, mutation (e.g., substitution) of at least one consensus nucleotide with a different, preferably non-consensus, nucleotide. These consensus nucleotides can be located within a consensus or near consensus splice site, or within an associated branch sequence (e.g., located upstream of a 3' splice acceptor site). Preferred consensus nucleotides for correction include invariant (i.e., conserved) nucleotides, including one or both of the invariant bases (AG) present in a 3' splice acceptor site; one or both of the invariant bases (GT) present in a 5' splice donor site; or the invariant A present in the branch sequence of a 3' splice acceptor site.

If the consensus or near consensus splice site is located within the coding region of a gene, then the correction is preferably achieved by conservative mutation. In a particularly preferred embodiment, all possible conservative mutations are made within a given consensus or near consensus splice site, so that the consensus or near consensus splice site is as far from

consensus as possible (i.e., has the least homology to consensus as is possible) without changing the coding sequence of the consensus or near consensus splice site.

In another embodiment, a novel DNA of the invention comprises at least one non-naturally occurring intron, either within a coding sequence or within a 5' and/or 3' non-coding sequence of the DNA. Novel DNAs comprising one or more non-naturally occurring introns may further comprise one or more consensus or near consensus splice sites which have been corrected as previously summarized.

In a particular embodiment of the invention, the present invention provides a novel gene encoding a human Factor VIII protein. This novel gene comprises one or more non-naturally occurring introns which serve to increase transcription of the gene, or to alter splicing of the gene. The gene may alternatively or additionally comprise one or more consensus splice sites or near consensus splice sites which have been corrected, also to increase transcription of the gene, or to alter splicing of the gene. In one embodiment, the Factor VIII gene comprises the coding region of the full-length human Factor VIII gene, except that the coding region has been modified to contain an intron spanning, overlapping or within the region of the gene encoding the β-domain. This novel gene is therefore expressed as a β-domain deleted human Factor VIII protein, since all or a portion of the β-domain coding sequence (defined by an intron) is spliced out during transcription.

A particular novel human Factor VIII gene of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. Another particular novel human Factor VIII gene of the invention comprises the coding region of the nucleotide sequence shown in SEQ ID NO:3 (nucleotides 1006-8237). Particular novel expression vectors of the invention comprise the complete nucleotide sequences shown in SEQ ID NOS: 2, 3 and 4. These vectors include novel 5' untranslated regulatory regions designed to provide high liver-specific expression of human Factor VIII protein.

In still other embodiments, the invention provides a method of increasing expression of a DNA sequence (e.g., a gene, such as a human Factor VIII gene), and a method of increasing the amount of mRNA which accumulates in the cytoplasm following transcription of a DNA sequence. In addition, the invention provides a method of altering the transcription pattern (e.g., splicing) of a DNA sequence. The methods of the present invention each involve correcting one or more consensus or near consensus splice sites within the nucleotide sequence of a DNA, and/or adding one or more non-naturally occurring introns into the nucleotide sequence of a DNA.

In a particular embodiment, the invention provides a method of simultaneously increasing expression of a gene encoding human Factor VIII protein, while also altering the gene's splicing pattern. The method involves inserting into the coding region of the gene an intron which spans, overlaps or is contained within the portion of the gene encoding the β-domain. The method may additionally or alternatively comprise correcting within either the

coding sequence or the 5' or 3' untranslated regions of the novel Factor VIII gene, one or more consensus or near consensus splice sites.

In yet another embodiment, the invention provides a method of producing a human Factor VIII protein, such as a β-domain deleted Factor VIII protein, by introducing an expression vector containing a novel human Factor VIII gene of the invention into a host cell capable of expressing the vector, under conditions appropriate for expression, and allowing for expression of the vector to occur.

Brief Description of the Figures

Figure 1 shows the nucleotide sequence of an RNA intron. The GU of the 5' splice donor site, the AG of the 3' splice acceptor site, and the A of the Branch are invariant bases (100% conserved and essential for recognition as splice sites). U is T in a DNA intron. The Branch sequence is located upstream from the 3' splice acceptor site at a distance sufficient to allow for lariat formation during spliceosome assembly (typically within 30-60 nucleotides). N is any nucleotide. Splicing will occur 5' of the GT base pair within the 5' splice donor site, and 3' of the AG base pair.

Figure 2 shows the conservative correction of a near consensus 3' splice acceptor site. The correction is made by silently mutating the A of the invariant (conserved) AG base pair to C, G, or T which does not affect the coding sequence of the intron because Ser is encoded by three alternate codons.

Figure 3 is a map of the coding region of a β-domain deleted human Factor VIII cDNA, showing the positions of the 99 silent point mutations which were made within the coding region (contained in plasmid pDJC) to conservatively correct all near consensus splice sites. Numbering of nucleotides begins with the ATG start coding of the coding sequence. Arrows above the map show positions mutated within near consensus 5' splice donor sites. Arrows below the map show positions mutated within near consensus 3' splice acceptor sites. Each "B" shown on the map shows a position mutated within a consensus branch sequence.

Figure 4A-4C shows the silent nucleotide substitution made at each of the 99 positions maked by arrows in Figure 3, as well as the codon containing the substitution and the amino acid encoded.

Figure 5A-5O is a comparison of the coding sequence of (a) plasmid pDJC (top) containing the coding region of the human β-domain deleted Factor VIII cDNA modified by making 99 conservative point mutations to correct all near consensus splice sites within the

coding region, and (b) plasmid p25D (bottom) containing the same coding sequence prior to making the 99 point mutations. Point mutations (substitions) are indicated by a "v" between the two aligned sequences and correspond to the positions within the pDJC coding sequence shown in Figure 3. Plasmid p25D contains the same coding region as does plasmid pCY-2 shown in Figure 7 and referred to throughout the text.

Figure 6 shows a map of plasmid pDJC including restriction sites used for cloning, regulatory elements within the 5' untranslated region, and the corrected human β-domain deleted Factor VIII cDNA coding sequence.

Figure 7 shows a map of plasmid pCY-2 including restriction sites used for cloning, regulatory elements within the 5' untranslated region, and the uncorrected (i.e., naturally-occurring) human β-domain deleted Factor VIII cDNA coding sequence. pCY-2 and pDJC are identical except for their coding sequences.

Figure 8 is a map of the human β-domain deleted Factor VIII cDNA coding region showing the five sections of the cDNA (delineated by restriction sites) which can be synthesized (using overlapping 60-mer oligonucleotides) to contain corrected near consensus splice sites, and then and assembled together to produce a new, corrected coding region.

Figure 9 is a schematic illustration of the cloning procedure used to insert an engineered intron into the coding region of the human Factor VIII cDNA, spanning a majority of the region of the cDNA encoding the β-domain. PCR fragments were generated containing nucleotide sequences necessary to create consensus 5' splice donor and 3' splice acceptor sites when cloned into selected positions flanking the β-domain coding sequence. The fragments were then cloned into plasmid pBluescript and sequenced. Once sequences had been confirmed, the fragments creating the 5' splice donor (SD) site were cloned into plasmid pCY-601 and pCY-6 (containing the full-length human Factor VIII cDNA coding region) immediatedly upstream of the β-domain coding sequence, and fragments creating the 3' splice acceptor (SA) site were cloned into pCY-601 and pCY-6 immediately downstream of the β-domain coding sequence. The resulting plasmids are referred to as pLZ-601 and pLZ-6, respectively.

Figure 10 is a map of the full-length human Factor VIII gene, showing the A1, A2, B, A3, C1 and C2 domains. Following expression of the gene, the β domain is naturally cleaved out of the protein. The map shows the 5' and 3' splice sites inserted within the B region of the gene (in plasmid pLZ-6) so that, during pre-mRNA processing of the gene, the majority

of the B region will be spliced out. Segments A2 and A3 of the gene will then be juxtaposed, coding for amino acids SFSQNPPV at the juncture.

Figure 11 shows the nucleotide sequences of the exon/intron boundaries (SEQ ID NO:5) flanking the β-domain coding region in plasmid pLZ-6 (containing the full-length human Factor VIII cDNA). The 5' splice donor site was added so that splicing would occur 5' of the "g" shown at position 2290. The 3' splice acceptor site was added so that splicing would occur 3' of the "g" shown at position 5147. Following splicing of the intron created by these splice sites, amino acids Gln-744 and Asn-1639 of the full-length human Factor VIII protein are brought together, resulting in a deletion of amino acids 745 to 1638 (numbering is in reference to Ala-1 of the mature human Factor VIII protein following cleavage of the 19 amino acid signal peptide). Capital letters represent nucleotide bases which remain within exons of the mRNA. Small case letters represent nucleotide bases which are spliced out of the mRNA as part of the intron.

Figure 12 is a map of the coding region of the full-length human Factor VIII gene showing (a) ATG (start) and TGA (stop) codons, (b) restriction sites within the coding region, (c) 5' splice donor (SD) and 3' splice acceptor (SA) sites of a rabbit β-globin intron positioned upstream of the coding region within the 5' untranslated region, (d) 5' splice donor and 3' splice acceptor sites added within the coding region defining an internal intron spanning the β-domain.

Figure 13 is a schematic illustration comparing the process of transcription, expression and post-translational modification for human Factor VIII produced from (a) a full-length human Factor VIII gene, (b) a β-domain deleted human Factor VIII gene, and (c) a full-length human Factor VIII gene containing an intron spanning the β-domain coding region.

Figure 14 is a graphic comparison of human Factor VIII expression for (a) pCY-6 (containing the coding region of the full-length human Factor VIII cDNA, as well as a 5' untranslated region derived from the second IVS of rabbit beta globin gene), (b) pCY-601 (containing the coding region of the full-length human Factor VIII cDNA, without the rabbit beta globin IVS), (c) pLZ-6 (containing the coding region of a full-length human Factor VIII cDNA with an intron spanning the β-domain, as well as the rabbit beta globin IVS), and (d) pLZ-601 (containing the coding region of a full-length human Factor VIII cDNA with an intron spanning the majority of the β-domain, without the rabbit beta globin IVS). Expression is given in nanograms. Transfection efficiencies were normalized to expression

of human growth hormone (hGH). Each bar represents a summary of four separate transfection experiments.

Figure 15 shows areas within the human Factor VIII transcription unit for sequence optimization.

Figure 16 shows the optimized intron-split leader sequence within vectors pCY-2, pCY-6, PLZ-6 and pCY2-SRE5, as well as the secondary structure of the leader sequence (SEQ ID NO:11) predicted by the computer program RNAdraw™.

Figure 17 is a schematic illustration showing two different RNA export pathways. The majority of mRNA's in higher eukaryotes contain intronic sequences which are removed within the nucleus (splicing pathway), follwed by export of the mRNA into the cytoplasm. Mammalian intronless genes, hepadnaviruses (e.g., HBV), and many retroviruses access a nonsplicing pathway which is facilitated by cellular RNA export proteins (facilitated pathway).

Figure 18 is a graph showing the effect of a 5' intron and 3' post-transcriptional regulatory element (PRE) on human Factor VIII expression levels in HuH-7 cells. Plasmid pCY-2 contains a 5' intron but no PRE. Plasmid pCY-201 is identical to pCY-2, except that it lacks the 5' intron. Plasmid pCY-401 and pCY-402 are identical to pCY-201, except that they contain one and two copies of the PRE, respectively. The levels of secreted active Factor VIII was measured from supernatants collected 48 hours (first bar of each group) or 72 hours (second bar of each group) after transfection by Coatest VIII: c/4 kit from Kabi Inc. The transfection efficiency of each plasmid was normalized by analysis of human growth hormone secreted levels.

Figure 19 is a graph comparing human Factor VIII expression *in vivo* in mice for plasmids containing various regulatory elements upstream of either the β-domain deleted or full-length human Factor VIII gene. Plasmid pCY-2 has a 5' untranslated region containing the liver-specific thyroxin binding globulin (TBG) promoter, two copies of the liver-specific alpha-1 microglobulin/bikunin (ABP) enhancer; and a modified rabbit β-globin IVS, all upstream of the human β-domain deleted Factor VIII gene. Plasmid pCY2-SE5 is identical to pCY-2 except that the TBG promoter was replaced by the endothelium-specific human endothelin-1 (ET-1) gene promoter, and the ABP enhancers (both copies) were replaced by one copy of the human c-fos gene (SRE) enhancer. Plasmid pCY-6 is identical to pCY-2, except that the human β-domain deleted Factor VIII gene was replaced by the full-length human Factor VIII gene. Plasmid pLZ-6 is identical to pCY-6, except that the full-length

human Factor VIII gene contained an intron spanning the β-domain. Plasmid pLZ-6A is identical to pLZ-6, except that it contains one corrected near consensus 3' splice acceptor site (A to C at base 3084 of pCY-6 (SEQ ID NO:3). Each bar represents an average of five mice.

Figure 20 shows the nucleotide sequence of the human alpha-1 microglobulin/bikunin (ABP) enhancer. Clustered liver-specific elements are underlined and labeled HNF-1, HNF-3 and HNF-4.

Figure 21 shows the nucleotide sequence of the human thyroxin binding globulin (TBG) promoter, also containing clustered liver-specific enhancer elements.

Figure 22 shows the nucleotide sequence and secondary structure of an optimized leader sequence.

Figure 23 is a comparison of the nucleotide sequences of the rabbit β-globin IVS before (top line) and after (bottom line) optimization to contain consensus 5' splice donor, 3' splice acceptor, branch, and translation initiation sites. Five nucleotides were also changed from purines to pyrimidines to optimize the pyrimidine track.

Figure 24 contains a list of various endothelium-specific promoters and enhancers, and characteristics associated with these promoters and enhancers.

Figure 25 is a graph comparing expression of plasmid pCY-2 and p25D *in vivo* in mice. Both plasmids contain the same coding sequence (for human β-domain deleted Factor VIII). Plasmid pCY-2 has an optimized 5' UTR containing two copies of the ABP enhancer, one copy of the TBG promoter and a leader sequence split by an optimized 5' rabbit β-globin intron. Plasmid p25D has a 5' UTR containing one copy of the CMV enhancer, one copy of the CMV promoter, and a leader sequence containing a short (130 bp) chimeric human IgE intron. Each bar represents an average of 5 mice.

Detailed Description of the Invention

DEFINITIONS

The present invention is described herein using the following terms which shall be understood to have the following meanings:

An "isolated DNA" means a DNA molecule removed from its natural sequence context (i.e., from its natural genome). The isolated DNA can be any DNA which is capable of being transcribed in a cell, including for example, a cloned gene (genomic or cDNA clone)

encoding a protein of interest, operably linked to a promoter. Alternatively, the isolated DNA can encode an antisense RNA.

A "5' consensus splice site" means a nucleotide sequence comprising the following bases: MAGGTRAGT, wherein M is (C or A), wherein R is (A or G) and wherein GT is essential for recognition as a 5' splice site (hereafter referred to as the "essential GT pair" or the "invariant GT pair").

A "3' consensus splice site" means a nucleotide sequence comprising the following bases (Y>8)NYAGG, wherein Y>8 is a pyrimidine track containing at least eight (most commonly twelve to fifteen or more) tandem pyrimidines (i.e., C or T (U if RNA)), wherein N comprises any nucleotide, wherein Y is a is a pyrimidine, and wherein the AG is essential for recognition as a 3' splice site (hereafter referred to as the "essential AG pair" or the "invariant AG pair"). A "3' consensus splice site" is also preceded upstream (at a sufficient distance to allow for lariat formation, typically at least about 40 bases) by a "branch sequence" comprising the following seven nucleotide bases: YNYTRAY, wherein Y is a pyrimidine (C or T), N is any nucleotide, R is a purine (A or G), and A is essential for recognition as a branch sequence (hereafter referred to as "the essential A" or the "invariant A"). When all seven branch nucleotides are located consecutively in a row, the branch sequence is a "consensus branch sequence."

A "near consensus splice site" means a nucleotide sequence which:

(a) comprises the essential 3' AT pair, and is at least about 50% homologous, more preferably at least about 60-70% homologous, and most preferably greater than 70% homologous to a 3' consensus splice site, when aligned with the consensus splice site for purposes of comparison; or

(b) comprises the essential 5' GT pair, and is at least about 50% homologous, more preferably at least about 60-70% homologous, and most preferably greater than 70% homologous to a 5' consensus splice site, when aligned with the consensus splice site for purposes of comparison.

Homology refers to sequence similarity between two nucleic acids. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences.

As will be described in more detail below, additional criteria for selecting "near consensus splice sites" can be used, adding to the definition provided above. For example, if a near consensus splice site shares homology with a 5' consensus splice site in only 5 out of 9 bases (i.e., about 55% homology), then these bases can be required to be located consecutively in a row. It can additionally or alternatively be required that a 3' near

consensus splice site be preceded by a consensus branch sequence (i.e., no mismatches allowed), or followed downstream by a consensus or near consensus 5' splice donor site, to make the selection more stringent.

The term "corrected" as used herein refers to a near consensus splice site mutated by substitution of at least one nucleotide shared with a consensus splice site, hereafter referred to as a "consensus nucleotide". The consensus nucleotide within the near consensus splice site is substituted with a different, preferably non-consensus nucleotide. This makes the near consensus splice site "farther from consensus."

If the near consensus splice site is within a coding region of a gene, then the correction is preferably a conservative mutation. A "conservative mutation" means a base mutation which does not affect the amino acid sequence coded for, also known as a "silent mutation." Accordingly, in a preferred embodiment of the invention, correction of a near consensus splice site located within the coding region of a gene includes making all possible conservative mutations to consensus nucleotides within the site, so that the near consensus splice site is as far from consensus as possible without changing the amino acid sequence it encodes.

A "Factor VIII gene" as used herein means a gene (e.g., a cloned genomic gene or a cDNA) encoding a functional human Factor VIII protein from any species (e.g., human or mouse). A Factor VIII gene which is "full-length" comprises the complete coding sequence of the human Factor VIII gene found in nature, including the region encoding the β-domain. A Factor VIII gene which "encodes a β-domain deleted Factor VIII protein" or "a β-domain deleted Factor VIII gene" lacks all or a portion of the region of the full-length gene encoding the β-domain and, therefore, is transcribed and expressed as a "truncated" or "β-domain deleted" Factor VIII protein. A gene which "is expressed as a β-domain deleted Factor VIII protein" includes not only a gene which encodes a β-domain deleted Factor VIII protein, but also a novel Factor VIII gene provided by the present invention which comprises the coding region of a full-length Factor VIII gene, except that it additionally contains an intron spanning the portion of the gene encoding the β-domain. The term "spans" means that the intron overlaps, encompasses, or is encompassed by the portion of the gene encoding the β domain. The portion of the gene spanned by the intron is then spliced out of the gene during transcription, so that the resulting mRNA is expressed as a truncated or β-domain deleted Factor VIII protein.

A "truncated" or "β-domain deleted" Factor VIII protein includes any active Factor VIII protein (human or otherwise) which contains a deletion of all or a portion of the β-domain..

A "non-naturally occurring intron" means an intron (defined by a 5' splice donor site and a 3' splice acceptor site) which has been engineered into a gene, and which is not present in the natural DNA or pre-mRNA nucleotide sequences of the gene.

An "expression vector" means any DNA vector (e.g., a plasmid vector) containing the necessary genetic elements for expression of a novel gene of the present invention. These elements, including a suitable promoter and preferably also a suitable enhancer, are "operably linked" to the gene, meaning that they are located at a position within the vector which enables them to have a functional effect on transcription of the gene.

IDENTIFICATION OF CONSENSUS AND NEAR CONSENSUS SPLICE SITES

A consensus or near consensus splice site can be identified within a DNA, or its corresponding RNA transcript, by evaluating the nucleotide sequence of the DNA for the presence of a sequence which is identical or highly homologous to either a 3' consensus splice acceptor site or a 5' consensus splice donor site (Figure 1). Such consensus and near consensus sites can be located within any portion of a given DNA (e.g., a gene), including the coding region of the DNA and any 3' and 5' untranslated regions.

To identify 3' consensus and near consensus splice acceptor sites, a DNA (or corresponding RNA) sequence is analyzed for the presence of one or more nucleotide sequences which includes an AG base pair, and which is either identical to or at least about 50% homologous, more preferably at least about 60-70% sequence homologous, to the sequence: **(T/C)≥8 N(C/T)AGG**. In a preferred embodiment, the nucleotide sequence is also followed upstream, typically by about 40 bases, by a nucleotide sequence which is identical to or highly homologous (e.g., at least about 50%-95% homologous) to a branch consensus sequence comprising the following bases: **(C/T)N(C/T)T(A/G)A(C/T)**, wherein N is any nucleotide, and A is invariant (i.e., essential). By way of example, in studies described herein, consensus and near consensus 3' splice sites were selected for correction within a gene encoding Factor VIII using the following criteria: the consensus or near consensus site (a) contained an AG pair, and (b) contained no more than three mismatches to a 3' consensus site.

To identify 5' consensus and near consensus splice donor sites, a DNA (or corresponding RNA) sequence can be analyzed for the presence of one or more nucleotide sequences which contains a GT base pair, and which is either identical to or at least about 50% homologous, more preferably at least about 60-70% homologous, to the sequence: **(A/C)AGGT(A/G)AGT**. By way of example, in studies described herein, consensus and near consensus 5' splice sites were selected for correction within a gene encoding Factor VIII using the following criteria: the consensus or near consensus site (a) contained a GT pair, and (b) contained no more than four mismatches to a 5' consensus site, provided that if it contained four mismatches, they were located consecutively in a row.

Evaluation of DNA or RNA sequences for the presence of one or more consensus or near consensus splice sites can be performed in any suitable manner. For example, nucleotide sequences can be manually analyzed. Alternatively, a computer algorithm can be

employed to search nucleotide sequences for specified base patterns (e.g., the MacVector™ program). The latter approach is preferred for large DNAs or RNAs, particularly because it allows for easy implementation of multiple search parameters.

CORRECTION OF CONSENSUS AND NEAR CONSENSUS SPLICE SITES

In one embodiment of the invention, splice and branch sequences which are consensus, or near consensus, are corrected by substitution of one or more consensus nucleotides within the site. The consensus nucleotide within the site is preferably substituted with a non-consensus nucleotide. For example, if the nucleotide being substituted is a C (i.e., a pyrimidine) and the consensus sequence contains either C or T, then the nucleotide is preferably substituted by an A or G (i.e., a purine), thereby making the consensus or near consensus splice site "farther from consensus."

In a preferred embodiment of the invention, consensus and near consensus sites which are located within a coding region of a gene are corrected by conservative substitution of one or more nucleotides so that the correction does not affect the amino acid sequence coded for. Such conservative or "silent" mutation of codons to preserve coding sequences is well known in the art. Accordingly, the skilled artisan will be able to select appropriate base substitutions to retain the coding sequence of any codon which forms all or part of a consensus or near consensus splice site. For example, as shown in Figure 2, if a 3' near consensus splice site contains a TCA codon encoding serine, and the A is a consensus nucleotide (e.g., part of the essential AG pair, then this nucleotide can be substituted with a C, G, or a T to correct the 3' near consensus splice site (e.g., making it no longer near consensus because it does not contain the essential AG pair required for a 3' near consensus splice site), without affecting the coding sequence of the codon.

Accordingly, in a preferred embodiment of the invention, correction of consensus or near consensus splice sites which are specifically located within the coding region of a gene is achieved by substitution of one or both bases of an essential AG or GT pair within the consensus or near consensus splice site, with a base which does not alter the coding sequence of the site. Correction of consensus or near consensus branch sequences is similarly achieved by substitution of the essential A within the consensus or near consensus branch site, with a base which does not alter the coding sequence of the site. By correcting any of these essential bases, the splice or branch site will no longer be consensus or near consensus.

In another preferred embodiment, correction of consensus or near consensus splice sites which are specifically located within the coding region of a gene is achieved by making all possible conservative mutations to consensus nucleotides within the site, so that the consensus or near consensus splice site is as far from consensus as possible but encodes the same amino acid sequence.

Other preferred corrections of the invention include corrections of 3' consensus and near consensus splice sites which are followed downstream (e.g., by approximately 50-350 nucleotides) by a consensus or near consensus 5' splice donor site. Other preferred corrections of the invention include corrections of 5' consensus and near consensus splice sites which are preceded upstream (e.g., by about 50-350 nucleotides) by a consensus or near consensus 3' splice acceptor site.

For consensus or near consensus splice sites which are located outside the coding region of a gene, for example, in a 3' or 5' untranslated region (UTR), alternative approaches to correction can also be employed. For instance, because preservation of the coding sequence is not a consideration, the near consensus splice site can be corrected not only by any base substitution, but also by addition or deletion of one or more bases within the consensus or near consensus splice site, making the site farther from consensus.

Techniques for making nucleotide base substitutions, additions and deletions as described above are well known in the art. For example, standard point mutation may be employed to substitute one or more bases within a near consensus splice site with a different (e.g., non-consensus) base. Alternatively, as described in detail in the examples below, entire genes or portions thereof can be reconstructed (e.g., resynthesized using PCR), to correct multiple consensus and near consensus splice sites within a particular region of a gene. This approach is particularly advantageous if a gene contains a high concentration of consensus and/or near consensus splice sites within a given region.

In a specific embodiment, the invention features a novel Factor VIII gene containing one or more consensus or near consensus splice sites which have been corrected by substitution of one or more consensus nucleotides within the site. As part of the present invention, the coding region of a gene (cDNA) encoding human β-domain deleted Factor VIII protein (nucleotides 1006-5379 of SEQ ID NO:2) was evaluated as described herein and found to contain 23 near consensus 5' splice (donor) sequences, 22 near consensus 3' splice (acceptor) sequences, and 18 consensus branch sequences (shown in Figure 3). A new coding sequence (SEQ ID NO:1) was then developed for this gene to correct all 3' and 5' near consensus splice sites by conservative mutation. In total, 99 point mutations were made to the coding region. The location of each of these point mutations is shown in Figure 3. The specific base substitution made in each of these point mutations is shown in Figure 4(A-C).

A comparison of this new coding sequence (SEQ ID NO:1) and the original uncorrected sequence (nucleotides 1006-5379 of SEQ ID NO:2), also showing the positions and specific substitutions made in each of the ninety-nine point mutations, is shown in Figure 5(A-O). A plasmid vector, referred to as pDJC, containing the new (i.e., corrected) Factor VIII gene coding sequence, including restriction sites used to synthesize the gene and regulatory elements used to express the gene, is shown in Figure 6. A plasmid vector,

referred to as pCY2, containing the original, uncorrected Factor VIII gene, including restriction sites and regulatory elements used to express the gene, is shown in Figure 7.

As described in further detail in the examples below, all 99 consensus base corrections within the coding region of pDJC can be made by synthesizing overlapping oligonucleotides (based on the sequence of pCY2 shown in SEQ ID NO:2) which contain the desired corrections. A schematic illustration of this process is shown in Figures 8. In total, 185 overlapping 60-mer oligonucleotides can be synthesized, and assembled in five segments using the method of Stemmer et al. (1995) *Gene* 164: 49-53. Prior to assembly, each segment can be sequenced and tested in *in vitro* transfection assays (e.g., nuclear and cytoplasmic RNA analysis) in pCY2.

As an alternative to the "correct all" approach described above, selective correction of consensus and near consensus splice sites can also be employed. This involves selecting only (a) consensus sites, and near consensus splice sites which are close to consensus, and/or (b) consensus sites and near consensus sites which are located at positions which render these sites more likely to function as a splice donor or acceptor site. To select only nucleotide sequences which are complete consensus or which are close to consensus, evaluation of a given nucleotide sequence is limited to analyzing the nucleotide sequence for sequences which are identical to or are highly homologous (e.g., greater than 70-80% homologous) to a 3' or 5' consensus splice site. To select only nucleotide sequences which are located at positions which render these sites more likely to function as a splice donor or acceptor site, the location of each 3' consensus or near consensus splice site must be evaluated with respect to the position of any neighboring 5' consensus or near consensus splice sites. If a 3' consensus or near consensus splice site is located approximately 50-350 bases upstream from a 5' consensus or near consensus splice site, then these 3' and 5' splice sites are likely to function as a splice acceptor and donor sites. Therefore, these sites are preferably, and selectively, removed.

By way of example, particular consensus and/or near consensus 5' splice donor and 3' splice acceptor sites, as shown in Figure 3, can be selected within the coding region of the cDNA encoding human β-domain deleted Factor VIII (nucleotides 1006-5379 of SEQ ID NO:2) for preferred correction, based on their relative locations (i.e., 3' splice acceptor site located approximately 50-350 bases upstream from 5' near consensus splice site). Such preferred selective corrections can include, for instance, the near consensus 3' splice acceptor site spanning nucleotide base 1851 of the coding region (see Figure 3) and any of the near consensus 5' splice donor sites located within 50-350 bases downstream of this near consensus 3' splice acceptor site, such as those spanning positions 1956, 1959, 2115, 2178 and 2184.

Splice site correction as provided herein can be applied to any gene known in the art. For example, the complete nucleotide sequence of other (e.g., full-length and β-domain

deleted) Factor VIII genes (both genomic clones and cDNAs) are described in US Patent No. 4,757,006, US Patent No. 5,618,789, US Patent No. 5,683,905, and US Patent No. 4,868,112, the disclosures of which are incorporated by reference herein. The nucleotide sequences of these genes can be analyzed for consensus and near consensus splice sites, and thereafter corrected, using the guidelines and procedures provided herein.

In addition, other genes, particularly large genes containing several introns and exons, are also suitable candidates for splice site correction. Such genes, include, for example, the gene encoding Factor IX, or the cystic fibrosis transmembrane regulator (CFTR) gene described in US Patent No. 5,240,846, or nucleic acids encoding CFTR monomers, as described in US Patent No. 5,639,661. The disclosures of both of these patents are accordingly incorporated by reference herein.

ADDITION OF INTRONS

In another embodiment, a novel gene of the invention includes one or more non-naturally occurring introns which have been added to the gene to increase expression of the gene, or to alter the splicing pattern of the gene. The present invention provides the first known instance of gene engineering which involved adding a non-naturally-occurring intron within the coding sequence of a gene, particularly without affecting the activity of the protein encoded by the gene. The benefit of intron addition in this context is at least two-fold. First, as shown in Figure 14 in the context of the human Factor VIII gene, addition of one or more introns into a gene increases the expression of the gene compared to the same gene without the intron. Second, the intron, when placed within the coding sequence of the gene, can be used to beneficially alter the splicing pattern of the gene (e.g., so that a particular protein of interest is expressed), and/or to increase cytoplasmic accumulation of mRNA transcribed from the gene.

Novel genes of the present invention may also contain introns outside of the coding region of the gene. For example, introns may be added to the 3' or 5' non-coding regions of the gene (utranslated regions (UTRs)). In a preferred embodiment of the invention, an intron is added upstream of the gene in the 5' UTR, as shown in pDJC (Figure 6) and pCY2 (Figure 7). Such introns may include newly engineered introns or pre-existing introns. In a preferred embodiment of the invention, the intron is derived from the rabbit β-globin intron (IVS).

In a particular embodiment, the invention provides a novel human Factor VIII gene which includes within its coding region one or more introns. If the gene comprises the coding region of a full-length human Factor VIII gene, then at least one of these introns preferably spans (i.e., overlaps, encompasses or is encompassed by) the portion of the gene encoding the β-domain. This portion of the gene is then spliced out during transcription of the gene, so that the gene is expressed as a β-domain deleted protein (i.e., a Factor VIII protein lacking all or a portion of the β-domain).

A β-domain deleted human Factor VIII protein possesses known advantages over a full-length human Factor VIII protein (also known as human Factor VIII:C), including reduced immunogenicity (Toole et al. (1986) *PNAS* 83: 5939-5942). Moreover, it is well known that the β-domain is not needed for activity of the Factor VIII protein. Thus, a novel Factor VIII gene of the invention provides the dual benefit of (1) increased and (2) preferred protein expression.

Addition of one or more introns into a gene can be achieved by adding a 5' splice donor site and a 3' splice acceptor site (Figure 1) into the nucleotide sequence of the gene at a desired location. If the intron is being added to remove a portion of the coding sequence from the gene, then a 5' splice donor site is placed at the 5' end of the portion being removed (i.e., defined by the intron) and a 3' splice acceptor site is placed at the 3' end of the portion to be removed. Preferably, the 5' splice donor and 3' splice acceptor sequences are consensus, including the branch sequence located upstream of the 3' splice site, so that they will be favored (and more likely bound) by cellular splicing machinery over any surrounding near consensus splice sites.

As shown in Figure 1, splicing will occur 5' of the essential GT base pair within the 5' splice donor site, and 3' of the essential AG base pair within the 3' splice acceptor site. Thus, for introns added to coding sequences of genes, the intron is preferably designed to that, upon splicing, the coding sequence is unaffected. This can be done by designing and adding 5' splice donor and 3' splice acceptor sites which include only conservative (i.e., silent) changes to the nucleotide sequence of the gene, so that addition of these splice sites does not alter the coding sequence.

For example, as part of the present invention, an intron was engineered into the coding sequence of a full-length cDNA encoding human Factor VIII (1006-8061 of SEQ ID NO:4). The intron spanned the portion of the gene encoding the β-domain (nucleotides 2290-5147 of SEQ ID NO:4, encoding amino acid residues 745-1638). As described in the examples below, this intron was created by adding a 5' splice donor site (100% consensus) so that splicing would occur immediately 5' of the coding sequence of the β-domain. A 3' splice acceptor site was also added so that splicing would occur immediately 3' of the coding sequence of the β-domain. Figure 11 shows the nucleotide sequences (SEQ ID NO:5) of the precise boundaries of the resulting intron that was added.

The nucleotide sequence for the 5' splice donor site of the added intron was derived from the pre-existing splice donor sequence found at the 5' end of IVS (Intron) 13 of genomic Factor VIII. This intron precedes exon 14, the exon which contains the sequence coding for the β-domain. The inserted sequence also contained the first nine bases of IVS 13 following the splice donor sequence.

The sequence for the 3' splice acceptor site was derived from the pre-existing splice acceptor sequence found at the 3' end of IVS 14 of genomic Factor VIII. This intron follows

exon 14, the β-domain-containing exon. The inserted 3' splice acceptor site also contained 130 bases upstream of the splice acceptor in IVS 14. This upstream region contains at least two near-consensus branch sequences.

Thus, both the 3' and 5' engineered splice sites were designed to take advantage of pre-existing nucleotide sequences within the β-domain region of the human Factor VIII gene.

The 5' splice donor, 3' splice acceptor, and branch sequences of the added intron were further modified so that they were 100% consensus (i.e., congruent to their respective consensus splicing sequences). Modifications (e.g., base substitutions) were chosen so as to not alter the coding sequence of bases located upstream of the 5' splice site and downstream of the 3' splice site (i.e., flanking the boundaries of the intron). A map showing the various domains of the full-length Factor VIII gene, along with the 5' splice donor and 3' splice acceptor sites inserted into the gene, is shown in Figure 10. The complete nucleotide sequences of the intron boundaries (i.e., 5' splice donor and 3' splice acceptor) are shown in Figure 11 (SEQ ID NO:5). A map showing the location of the location of the 5' splice donor and 3' splice acceptor sites with respect to various restriction sites (used to clone in the sites) is shown in Figure 12. As shown schematically in Figure 13, the resulting novel Factor VIII gene, in contrast to a full-length Factor VIII gene or a gene encoding β-domain deleted Factor VIII, is transcribed as a pre-mRNA which contains the region encoding the β-domain, but is then spliced to remove the majority of this region, so that the resulting mRNA is expressed as a β-domain deleted protein. A complete expression plasmid (pLZ-6) containing the coding sequence of this novel Factor VIII gene, as well as an engineered 5' untranslated region containing regulatory elements designed to provide high, liver-specific expression, comprises the nucleotide sequence shown in SEQ ID NO:3. Bases 1006-8237 of pLZ-6 (SEQ ID NO:3) correspond to the coding region of the novel Factor VIII gene.

Accordingly, in a preferred embodiment, the invention provides a novel Factor VIII gene comprising a non-naturally occurring intron spanning all or a portion of the β-domain region of the gene. In one embodiment, the gene comprises the coding region of the nucleotide sequence shown in SEQ ID NO:3. The gene may also contain further modifications, such as additional introns, or one or more corrected consensus or near consensus splice sites as described herein. In particular, the gene may further comprise one or more introns upstream of the coding sequence of the gene, within the 5' UTR. As shown in Figures 6 and 7, a preferred intron for insertion within this region is the rabbit β-globin intron (IVS). In addition, consensus and near consensus splice site corrections can be made to the gene, such as those shown in Figures 3 and 4(A-C).

OPTIMIZATION OF 5' AND 3' UNTRANSLATED REGIONS FOR HIGH TISSUE-SPECIFIC GENE EXPRESSION

Novel DNAs of the invention are preferably in a form suitable for transcription and/or expression by a cell. Generally, the DNA is contained in an appropriate vector (e.g., an expression vector), such as a plasmid, and is operably linked to appropriate genetic regulatory elements which are functional in the cell. Such regulatory sequences include, for example, enhancer and promoter sequences which drive transcription of the gene. The gene may also include appropriate signal and polyadenylation sequences which provide for trafficking of the encoded protein to intracellular destinations or export of the mRNA. The signal sequence may be a natural sequence of the protein or an exogenous sequence.

Suitable DNA vectors are known in the art and include, for example, DNA plasmids and transposable genetic elements containing the aforementioned genetic regulatory and processing sequences. Particular expression vectors which can be used in the invention include, but are not limited to, pUC vectors (e.g., pUC19) (University of California, San Francisco) pBR322, and pcDNA1 (InVitrogen, Inc.). An expression plasmid, pMT2LA8, encoding a β-domain deleted Factor VIII protein is described, for example, by Pitman et al. (1993) *Blood* 81(11):2925-2935). Entire coding sequences for these plasmid vectors are also provided herein (SEQ ID NOS: 4 and 2, respectively).

Suitable regulatory sequences required for gene transcription, translation, processing and secretion are art-recognized, and are selected to direct expression of the desired protein in an appropriate cell. Accordingly, the term "regulatory sequence", as used herein, includes any genetic element present 5' (upstream) or 3' (downstream) of the translated region of a gene and which control or affect expression of the gene, such as enhancer and promoter sequences (e.g., viral promoters, such as SV40 and CMV promoters). Such regulatory sequences are discussed, for example, in Goeddel, Gene expression Technology: Methods in Enzymology, page 185, Academic Press, San Diego, CA (1990), and can be selected by those of ordinary skill in the art for use in the present invention.

In a preferred embodiment of the invention, the 5' and/or 3' untranslated regions (UTRs) of a gene construct (e.g., a novel DNA of the invention) are optimized to provide high, tissue-specific expression. Such optimization can include, for example, selection of optimal tissue-specific promoters and enhancers, multerimization of genetic elements, insertion of one or more introns within or outside of the coding sequence, correction of near-consensus 5' splice donor and 3' splice acceptor sites within or outside of the coding sequence, optimization of transcription initiation and termination sites, insertion of RNA export elements, and addition of polyadenylation trimer cassettes to insulate transription. In preferred embodiments of the invention, a combination of the aforementioned elements and sequence modifications are selected and engineered into the gene construct to provide optimized expression.

For many applications of human gene therapy, it is desirable to express proteins in the liver, which has the highest rate of protein synthesis per gram of tissue. For example, effective gene therapy for human Factor VIII requires sufficient levels and duration of protein expression in hepatocytes where Factor VIII is naturally produced, and/or in endothelial cells (ECs) where von Willebrand factor is produced, a protein which stabilizes the secretion of Factor VIII. Thus, in one embodiment, the invention provides a gene construct (e.g., expression vector) optimized to produce high levels and duration of liver-specific protein expression. In a particular embodiment, the invention provides a human Factor VIII gene construct, optimized to produce high levels and duration of liver-specific or endothelium-specific protein expression. This is achieved, for example, by selecting optimal liver-specific and endothelium-specific promoters and enhancers, and by combining these tissue-specific elements with other genetic elements and modifications to increase gene transcription.

Accordingly, for high levels and duration of gene expression in the liver, suitable promoters include, for example, promoters known to contain liver-specific elements. In one embodiment, the invention employs the thyroid binding globulin (TBG) promoter described by Hayashi et al. (1993) *Molec. Endocrinol.* 7:1049-1060. As shown in Figure 21, the TBG promoter contains hepatic nuclear factor (HNF) enhancer elements and provides the additional advantage of having a precisely mapped transcriptional start site. This allows insertion of a leader sequence, preferably optimized as described herein, between the promoter and the transcriptional start site. Figure 21 also shows the complete nucleotide sequence of the TBG promoter (SEQ ID NO:10).

For high levels and duration of gene expression in endothelium, suitable endothelium-specific promoters include, for example, the human endothelin-1 (ET-1) gene promoter described by Lee et al. (1990) *J. Biol. Chem.* 265(18), the fms-like tyrosine kinase promoter (Flt-1) described by Morishita et al. (1995) *J. Biol. Chem.* 270(46), the Tie-2 promoter described by Korhonen et al. (1995) *Blood* 86(5):1828-1835, and the nitric oxide synthase promoter described by Zhang et al. (1995) *J. Biol. Chem.* 270(25)) (*see* Figure 24).

Promoters selected for use in the invention are preferably paired with a suitable ubiquitous or tissue-specific enhancer designed to augment transcription levels. For example, in one embodiment, a liver-specific promoter, such as the TBG promoter, is used in conjunction with a liver-specific enhancer. In a preferred embodiment, the invention employs one or more copies of the liver-specific alpha-1 microglobulin/bikunin (ABP) enhancer described by Rouet et al. (1992) *J. Biol. Chem.* 267:20765-20773, in combination with the TBG promoter. As shown in Figure 20, the ABP enhancer contains a cluster of HNF enhancer elements common to many liver-specific genes within a short nucleotide sequence, making it suitable to multerimize. When multerimized, the ABP enhancer generally exhibits increased activity and functions in either orientation within a gene construct.

Thus, in one embodiment, the invention provides an expression vector or DNA construct comprising one or more copies of a liver-specific or endothelium-specific promoter and a liver-specific or endothelium-specific enhancer, the promoter and enhancer being derived from different genes, such as thyroid binding globulin gene and the alpha-1 microglobulin/bikunin gene.

Alternatively, strong ubiquitous (i.e., non-tissue specific) enhancers can be used in conjunction with tissue-specific promoters, such as the TBG promoter or the ET-1 promoter, to achieve high levels and duration of tissue-specific expression. Such ubiquitous enhancers include, for example, the human c-fos (SRE) gene enhancer described by Treisman et al. (1986) *Cell* 46 which, when used in combination with liver-specific promoters (e.g., TBG) or endothelium-specific promoters (e.g., ET-1), provide high levels of tissue-specific expression, as demonstrated in studies described herein.

Accordingly, in a particular embodiment, the invention provides a gene construct which is optimized for specific expression in liver cells by inserting within its 5' untranslated region one or more copies of the ABP enhancer (preferably two copies) coupled upstream with the TBG promoter, as shown in Figure 15. Specific gene constructs, such as pCY2 and pDJC, containing these elements inserted upstream of the coding region for human Factor VIII (β-domain deleted and full-length with intron spanning the β-domain), are shown in Figures 6 and 7, respectively. In another particular embodiment, the gene construct is optimized for specific expression in endothelial cells by inserting within its 5' region one or more copies of the c-fos SRE enhancer, or an endothelial-specific enhancer (e.g., the human tissue factor (hTF/m) enhancer described by Parry et al. (1995) *Arterioscler. Thromb. Vasc. Biol.* 15:612-621) coupled upstream with the ET-1 promoter.

In addition to selecting optimal promoters and enhancers, optimization of a gene construct can include the use of other genetic elements within the transcriptional unit of the gene to increase and/or prolong expression. In one embodiment, one or more introns (e.g., non-naturally occurring introns) are inserted into the 5' or 3' untranslated region (UTR) of the gene. Introns from a broad variety of known genes (e.g., mammalian genes) can be used for this purpose. In one embodiment, the invention employs the first intron (IVS) from the rabbit β-globin gene comprising the nucleotide sequence shown in Figure 23 (SEQ ID NO:6).

In cases where the intron does not contain consensus 5' splice donor and 3' splice acceptor sites, or a consensus branch and pyrimidine track sequence, the intron is preferably optimized (modified) to render these sites completely consensus. This can be achieved, for example, by substituting one or more nucleotides within the 5' or 3' splice site, as previously described herein to render the site consensus. For example, when using the rabbit β-globin intron, the nucleotide sequence can be modified as shown in Figure 16 to render the 5' splice donor and 3' splice acceptor sites, and the pyrimidine track, entirely consensus. This can facilitate efficient transcription and export of the gene message out of the cell nucleus,

thereby increasing expression. Exemplary nucleotide substitutions within the rabbit β-globin IVS which can be made to achieve this result are shown in Figure 23 which shows a comparison of the sequence for the unmodified (wild-type) rabbit β-globin intron (SEQ ID NO:6) and the same sequence modified to render the 5' splice donor and 3' splice acceptor sites, and the pyrimidine track, entirely consensus (SEQ ID NO:7).

When engineering one or more introns into the 5' UTR of a gene construct, the intron can be inserted into the leader sequence of the gene, as shown in Figures 15, 16 and 22. Accordingly, the intron can be inserted within the leader sequence, downstream from the promoter and enhancer elements. This can be done in conjunction with one or more additional modifications to the leader sequence, all of which serve to increase transcription, stability and export of mRNAs. Such additional modifications include, for example, optimizing the translation initiation site (Kozak et al. (1986) *Cell* 44:283) and/or the secondary structure of the leader sequence (Kozak et al. (1994) *Molec. Biol.* 235:95).

Accordingly, in a preferred embodiment, the invention provides a gene construct which contains within its transcriptional unit, one or a combination of the foregoing genetic elements and sequence modifications designed to provide high levels and duration of gene expression, optionally in a tissue-specific manner. In a particular embodiment, the construct contains a gene encoding human Factor VIII (e.g., β-domain deleted or full-length), having a 5' untranslated region which is optimized to provide significant levels and duration of liver-specific or endothelium-specific expression.

Particularly preferred gene constructs of the invention include, for example, those comprising the nucleotide sequences shown in SEQ ID NO:2 and SEQ ID NO:4, referred to herein respectively as pCY-2 and pLZ-6. These constructs contain the coding sequences for human β-domain deleted Factor VIII (pCY-2) and full-length human Factor VIII (containing an intron spanning the β-domain) (pLZ-6) downstream from an optimized 5' UTR designed to provide high levels and duration of human Factor VIII expression in liver cells. Other preferred gene constructs comprise the identical 5' UTR of pCY-2 and pLZ-6, in conjunction with coding sequences for other proteins desired to be expressed in the liver (e.g., other blood coagulation factors, such as human Factor IX).

As shown in Figures 7, 15 and 16, plasmids pCY-2 and pLZ-6 contain 5' UTRs comprising a novel combination of regulatory elements and sequence modifications shown herein to provide high levels and duration of human Factor VIII expression, both *in vitro* and *in vivo*, in liver cells. Specifically, each construct comprises within its 5' UTR sequentially from 5' to 3' (a) two copies of the ABP enhancer (SEQ ID NO:9), (b) one copy of the TBG promoter (SEQ ID NO:10), and (c) an optimized 71 nucleotide leader sequence (SEQ ID NO: 11) split by intron 1 of the rabbit β-globin gene. The intron is optimized to contain consensus splice acceptor, donor and pyrimidine track sites.

The leader sequence within the 5' UTR of pCY-2 and pLZ-6 also contains an optimized translation initiation site (SEQ ID NO: 8). Specifically, the human Factor VIII gene contains a cytosine at the +4 position, following the AUG start codon. This base was changed to a guanine, resulting in an amino acid change within the signal sequence of the protein from a glutamine to a glutamic acid. The leader sequence was further designed to have no RNA secondary structure, as predetermined by an RNA-folding algorithm (Figure 16) (Kozak et al. (1994) *J. Mol. Biol.* 235:95).

In addition to optimization of the 5' UTR of a gene construct, the 3' UTR can also be engineered to include one or more genetic elements or sequence modifications which increase and/or prolong expression of the gene. For example, the 3' UTR can be modified to provide optimal RNA processing, export and mRNA stability. In one embodiment of the invention, this is done by increasing translational termination efficiency. In mammalian RNA's, translational termination is generally optimal if the base following the stop codon is a purine (McCaughan et al. (1995) *PNAS* 92:5431). In the case of the human Factor VIII gene, the UGA stop codon is followed by a guanine and is thus already optimal. However, in other gene constructs of the invention which do not naturally contain an optimized translational termination sequence, the termination sequence can be optimized using, for example, site directed mutagenesis, to substitute the base following the stop codon for a purine.

In particular gene constructs of the invention which contain the human Factor VIII gene, the 3' UTR can further be modified to remove one or more of the three pentamer sequences AUUUA present in the 3' UTR of the gene. This can increase the stability of the message. Alternatively, the 3' UTR of the human Factor VIII gene, or any gene having a short-lived messenger RNA, can be switched with the 3' UTR of a gene associated with a message having a longer lifespan.

Additional modifications for optimizing gene constructs of the invention include insertion of one or more poly A trimer cassettes for optimal polyadenylation and 3' end formation. These can be inserted within the 5' UTR or the 3' UTR of the gene. In a preferred embodiment, the gene construct is flanked on either side by a poly A trimer cassette, as shown in Figure 15. These cassettes can inhibit transcription originating outside of the desired promoter in the transcriptional unit, ensuring that transcription of the gene occurs only in the tissue where the promoter is active (Maxwell et al. (1989) *Biotechniques* 1989 3:276). Additionally, because the poly A trimer cassette functions in both orientations, i.e., on each DNA strand, it can be utilized at the 3' end of the gene for transcriptional termination and polyadenylation, as well as to inhibit bottom strand transcription and production of antisense RNA.

In further embodiments of the invention, gene optimization includes the addition of viral elements for accessing non-splicing RNA export pathways. The majority of mRNAs in higher eukaryotes contain intronic sequences which are removed within the nucleus, followed

by export of the mRNA into the cytoplasm. This is referred to as the splicing pathway. However, as shown in Figure 17, mammalian intronless genes, hepadnaviruses (e.g., HBV), and many retroviruses access a nonsplicing pathway which is facilitated by cellular RNA export proteins and/or specific sequences within. This is referred to as the facilitated pathway.

In a particular embodiment, the gene construct is modified to include one or more copies of the post-transcriptional regulatory element (PRE) from hepatitis B virus. This 587 base pair element and its function to facilitate export of mRNAs from the nucleus, is described in U.S. Patent No. 5,744,326. Generally, the PRE element is placed within the 3' UTR of the gene, and can be inserted as two or more copies to further increase expression, as shown in Figure 18 (plasmid pCY-401 verses plasmid pCY-402).

Gene constructs (e.g., expression vectors) of the invention can still further include sequence elements which impart both an autonomous replication activity (i.e., so that when the cell replicates, the plasmid replicates as well) and nuclear retention as an episome. Generally, these sequence elements are included outside of the transcriptional unit of the gene construct. Suitable sequences include those functional in mammalian cells, such as the oriP sequence and EBNA-1 gene from the Epstein-Barr virus (Yates et al. (1985) *Nature* 313:812). Other suitable sequences include the *E. coli* origen of replication, as shown in Figures 6 and 7.

Gene constructs of the invention, such as pDJC, pCY-2, pCY-6, pLZ-6 and pCY2-SE5, have been described above, but are not intended to be limiting. Other novel constructs can be made in accordance with the guidelines provided herein, and are intended to be included within the scope of the present invention.

INCREASED CYTOPLASMIC RNA ACCUMULATION AND EXPRESSION

Novel DNAs (e.g., genes) of the present invention are modified to increase expression, for example, by facilitate cytoplasmic accumulation of mRNA transcribed from the DNA and by optimizing the 5' and 3' untranslated regions of the DNA. Accordingly, cytoplasmic mRNA accumulation and/or expression of the DNA is increased relative to the same DNA in unmodified form.

To evaluate (e.g., quantify) levels of nuclear or cytoplasmic mRNA accumulation obtained following transcription of novel DNAs and vectors of the invention, a variety of art recognized techniques can be employed, such as those described in Sambrook et al. "Molecular Cloning," 2d ed., and in the examples below. Such techniques include, for instance, Northern blot analysis, using total nuclear or cytoplasmic RNA. This assay can, optionally, be normalized using mRNA transcribed from a control gene, such as a gene encoding glyceraldehyde phosphate dehydrogenase (GAPDH). Levels of nuclear and cytoplasmic RNA accumulation can then be compared for novel DNAs of the invention to determine whether an increase has occurred following correction of one or more consensus or

near consensus splice sites, and/or by addition of one or more non-naturally occurring introns into the DNA.

Novel DNAs of the invention can also be assayed for altered splicing patterns using similar techniques. For example, as described in the examples below, to determine whether a non-naturally occurring intron has been successfully incorporated into a DNA so that it is correctly spliced during mRNA processing, cytoplasmic mRNA can be assayed by Northern blot analysis, reverse transcriptase PCR (RT-PCR), or RNase protection assays. Such assays are used to determine the size of the mRNA produced from the novel DNA containing the non-naturally occurring intron. The size of the mRNA can then be compared to the size of the DNA with and without the intron to determine whether splicing has been achieved, and whether the splicing pattern corresponds to that expected based on the size of the added intron.

Alternatively, protein expressed from cytoplasmic RNA can be assayed by SDS-PAGE analysis and sequenced to confirm that correct splicing has been achieved.

To measure expression levels, novel DNAs of the invention can also be tested in a variety of art-recognized expression assays. Suitable expression assays, as illustrated in the examples provided below, include quantitative ELISA (Zatloukal et al. (1994) *PNAS* 91: 5148-5152), radioimmunoassay (RIA), and enzyme activity assays. When expression of Factor VIII protein is being measured, in particular, Factor VIII activity assays such as the KabiCoATest, (Kabi Inc., Sweden) can be employed to quantify expression.

GENE DELIVERY TO CELLS

Following insertion into an appropriate vector, novel DNAs of the invention can be delivered to cells either *in vitro* or *in vivo*. For example, the DNA can be transfected into cells *in vitro* using standard transfection techniques, such as calcium phosphate precipitation (O'Mahoney et al. (1994) *DNA & Cell Biol.* 13(12): 1227-1232). Alternatively, the gene can be delivered to cells *in vivo* by, for example, intravenous or intramuscular injection.

In one embodiment of the invention, the gene is targeted for delivery to a specific cell by linking the plasmid to a carrier molecule containing a ligand which binds to a component on the surface of a cell, thereby forming a polynucleotide-carrier complex. The carrier can further comprise a nucleic acid binding agent which noncovalently mediates linkage of the DNA to the ligand of the carrier molecule.

The carrier molecule of the polynucleotide-carrier complex performs at least two functions: (1) it binds the polynucleotide (e.g., the plasmid) in a manner which is sufficiently stable (either in *vivo, ex vivo,* or *in vitro*) to prevent significant uncoupling of the polynucleotide extracellularly prior to internalization by a target cell, and (2) it binds to a component on the surface of a target cell so that the polynucleotide-carrier complex is internalized by the cell. Generally, the carrier is made up of a cell-specific ligand and a

cationic moiety which, for example are conjugated. The cell-specific ligand binds to a cell surface component, such as a protein, polypeptide, carbohydrate, lipid or combination thereof. It typically binds to a cell surface receptor. The cationic moiety binds, e.g., electrostatically, to the polynucleotide.

The ligand of the carrier molecule can be any natural or synthetic ligand which binds a cell surface receptor. The ligand can be a protein, polypeptide, glycoprotein, glycopeptide, glycolipid or synthetic carbohydrate which has functional groups that are exposed sufficiently to be recognized by the cell surface component. It can also be a component of a biological organism such as a virus, cells (e.g., mammalian, bacterial, protozoan).

Alternatively, the ligand can comprise an antibody, antibody fragment (e.g., an $F(ab')_2$ fragment) or analogues thereof (e.g., single chain antibodies) which binds the cell surface component (see e.g., Chen et al. (1994) *FEBS Letters* 338:167-169, Ferkol et al. (1993) *J. Clin. Invest.* 92:2394-2400, and Rojanasakul et al. (1994) *Pharmaceutical Res.* 11(12):1731-1736). Such antibodies can be produced by standard procedures.

Ligands useful in forming the carrier will vary according to the particular cell to be targeted. For targeting hepatocytes, proteins, polypeptides and synthetic compounds containing galactose-terminal carbohydrates, such as carbohydrate trees obtained from natural glycoproteins or chemically synthesized, can be used. For example, natural glycoproteins that either contain terminal galactose residues or can be enzymatically treated to expose terminal galactose residues (e.g., by chemical or enzymatic desialylation) can be used. In one embodiment, the ligand is an asialoglycoprotein, such as asialoorosomucoid, asialofetuin or desialylated vesicular stomatitis virus. In another embodiment, the ligand is a tri- or tetra-antennary carbohydrate moiety.

Alternatively, suitable ligands for targeting hepatocytes can be prepared by chemically coupling galactose-terminal carbohydrates (e.g., galactose, mannose, lactose, arabinogalactan etc.) to nongalactose-bearing proteins or polypeptides (e.g., polycations) by, for example, reductive lactosamination. Methods of forming a broad variety of other synthetic glycoproteins having exposed terminal galactose residues, all of which can be used to target hepatocytes, are described, for example, by Chen et al. (1994) *Human Gene Therapy* 5:429-435 and Ferkol et al. (1993) *FASEB* 7: 1081-1091 (galactosylation of polycationic histones and albumins using EDC); Perales et al. (1994) *PNAS* 91:4086-4090 and Midoux et al. (1993) *Nucleic Acids Research* 21(4):871-878 (lactosylation and galactosylation of polylysine using α-D-galactopyranosyl phenylisothiocyanate and 4-isothiocyanatophenyl β-D-lactoside); Martinez-Fong (1994) *Hepatology* 20(6):1602-1608 (lactosylation of polylysine using sodium cyanoborohydride and preparation of asialofetuin-polylysine conjugates using SPDP); and Plank et al. (1992) *Bioconjugate Chem.* 3:533-539 (reductive coupling of four

terminal galactose residues to a synthetic carrier peptide, followed by linking the carrier to polylysine using SPDP).

For targeting the polynucleotide-carrier complex to other cell surface receptors, the carrier component of the complex can comprise other types of ligands. For example, mannose can be used to target macrophages (lymphoma) and Kupffer cells, mannose 6-phosphate glycoproteins can be used to target fibroblasts (fibro- sarcoma), intrinsic factor-vitamin B12 and bile acids (See Kramer et al. (1992) J. Biol. Chem. 267:18598- 18604) can be used to target enterocytes, insulin can be used to target fat cells and muscle cells (see e.g., Rosenkranz et al. (1992) *Experimental Cell Research* 199:323-329 and Huckett et al. (1990) *Chemical Pharmacology* 40(2):253-263), transferrin can be used to target smooth muscle cells (see e.g., Wagner et al. (1990) *PNAS* 87:3410-3414 and U.S. Patent No. 5, 354,844 (Beug et al.)), Apolipoprotein E can be used to target nerve cells, and pulmonary surfactants, such as Protein A, can be used to target epithelial cells (see e.g., Ross et al. (1995) *Human Gene Therapy* 6:31-40).

The cationic moiety of the carrier molecule can be any positively charged species capable of electrostatically binding to negatively charged polynucleotides. Preferred cationic moieties for use in the carrier are polycations, such as polylysine (e.g., poly-L-lysine), polyarginine, polyornithine, spermine, basic proteins such as histones (Chen et al., *supra.*), avidin, protamines (see e.g., Wagner et al., *supra.*), modified albumin (i.e., N-acylurea albumin) (see e.g., Huckett et al., *supra.*) and polyamidoamine cascade polymers (see e.g., Haensler et al. (1993) *Bioconjugate Chem.* 4: 372-379). A preferred polycation is polylysine (e.g., ranging from 3,800 to 60,000 daltons). Other preferred cationic moieties for use in the carrier are cationic liposomes.

In one embodiment, the carrier comprises polylysine having a molecular weight of about 17,000 daltons (purchased as the hydrogen bromide salt having a MW of a 26,000 daltons), corresponding to a chain length of approximately 100-120 lysine residues. In another embodiment, the carrier comprises a polycation having a molecular weight of about 2,600 daltons (purchased as the hydrogen bromide salt having a MW of a 4,000 daltons), corresponding to a chain length of approximately 15-10 lysine residues.

The carrier can be formed by linking a cationic moiety and a cell-specific ligand using standard cross-linking reagents which are well known in the art. The linkage is typically covalent. A preferred linkage is a peptide bond. This can be formed with a water soluble carbodiimide, such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC), as described by McKee et al (1994) Bioconjugate Chem. 5: 306-311 or Jung, G. et al. (1981) Biochem. Biophys. Res. Commun. 101: 599-606 or Grabarek et al. (1990) Anal. Biochem. 185:131. Alternative linkages are disulfide bonds which can be formed using cross-linking reagents, such as N-Succinimidyl 3-(2-pyridyldithio)propionate (SPDP), N-hydroxysuccinimidyl ester of chlorambucil, N-Succinimidyl-(4-Iodoacetyl)aminobenzoate)

(SIAB), Sulfo-SIAB, and Sulfo-succinimidyl-4-maleimidophenyl-butyrate (Sulfo-SMPB). Strong noncovalent linkages, such as avidin-biotin interactions, can also be used to link cationic moieties to a variety of cell binding agents to form suitable carrier molecules.

The linkage reaction can be optimized for the particular cationic moiety and cell binding agent used to form the carrier. The optimal ratio (w:w) of cationic moiety to cell binding agent can be determined empirically. This ratio will vary with the size of the cationic moiety (e.g., polycation) being used in the carrier, and with the size of the polynucleotide to be complexed. However, this ratio generally ranges from about 0.2-5.0 (cationic moiety : ligand). Uncoupled components and aggregates can be separated from the carrier by molecular sieve or ion exchange chromatography (e.g., Aquapore™ cation exchange, Rainin).

In one embodiment of the invention, a carrier made up of a conjugate of asialoorosomucoid and polylysine is formed with the cross linking agent 1-(3-dimethylaminopropyl)-3-ethyl carbodiimide. After dialysis, the conjugate can be separated from unconjugated components by preparative acid-urea polyacrylamide gel electrophoresis (pH 4-5).

Following formation of the carrier molecule, the polynucleotide (e.g., plasmid) is linked to the carrier so that (a) the polynucleotide is sufficiently stable (either in *vivo, ex vivo,* or *in vitro*) to prevent significant uncoupling of the polynucleotide extracellularly prior to internalization by the target cell, (b) the polynucleotide is released in functional form under appropriate conditions within the cell, (c) the polynucleotide is not damaged and (d) the carrier retains its capacity to bind to cells. Generally, the linkage between the carrier and the polynucleotide is noncovalent. Appropriate noncovalent bonds include, for example, electrostatic bonds, hydrogen bonds, hydrophobic bonds, anti-polynucleotide antibody binding, linkages mediated by intercalating agents, and streptavidin or avidin binding to polynucleotide-containing biotinylated nucleotides. However, the carrier can also be directly (e.g., covalently) linked to the polynucleotide using, for example, chemical cross-linking agents (e.g., as described in WO-A-91/04753 (Cetus Corp.), entitled "Conjugates of Antisense Oligonucleotides and Therapeutic Uses Thereof").

As described in Example 4, polynucleotide-carrier complexes can be formed by combining a solution containing carrier molecules with a solution containing a polynucleotide to be complexed, preferably so that the resulting composition is isotonic (see Example 4).

ADMINISTRATION

Novel DNAs of the invention can be administered to cells either *in vitro* or *in vivo* for transcription and/or expression therein.

For *in vitro* delivery, cultured cells can be incubated with the DNA in an appropriate medium under suitable transfection conditions, as is well known in the art.

For *in vivo* delivery (e.g., in methods of gene therapy) DNAs of the invention (preferably contained within a suitable expression vector) can be administered to a subject in a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier", as used herein, is intended to include any physiologically acceptable vehicle for stabilizing DNAs of the present invention for administration *in vivo*, including, for example, saline and aqueous buffer solutions, solvents, dispersion media, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media is incompatible with the polynucleotide-carrier complexes of the present invention, use thereof in a therapeutic composition is contemplated.

Accordingly, novel DNAs of the invention can be combined with pharmaceutically acceptable carriers to form a pharmaceutical composition. In all cases, the pharmaceutical composition must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action or microorganisms such as bacteria and fungi. Protection of the polynucleotide-carrier complexes from degradative enzymes (e.g., nucleases) can be achieved by including in the composition a protective coating or nuclease inhibitor. Prevention of the action of microorganisms can be achieved by various anti-bacterial and anti-fungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like.

Novel DNAs of the invention may be administered *in vivo* by any suitable route of administration. The appropriate dosage may vary according to the selected route of administration. The DNAs are preferably injected intravenously in solution containing a pharmaceutically acceptable carrier, as defined herein. Sterile injectable solutions can be prepared by incorporating the DNA in the required amount in an appropriate buffer with one or a combination of ingredients enumerated above or below, followed by filtered sterilization. Other suitable routes of administration include intravascular, subcutaneous (including slow-release implants), topical and oral.

Appropriate dosages may be determined empirically, as is routinely practiced in the art. For example, mice can be administered dosages of up to 1.0 mg of DNA per 20 g of mouse, or about 1.0 mL of DNA in solution per 1.4 mL of mouse blood.

Administration of a novel DNA, or protein expressed therefrom, to a subject can be in any pharmacological form including a therapeutically active amount of DNA or protein, in combination with another therapeutic molecule. Administration of a therapeutically active amount of a pharmaceutical composition of the present invention is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result (e.g., an improvement in clinical symptoms). A therapeutically active amount of DNA or

protein may vary according to factors such as the disease state, age, sex, and weight of the individual. Dosage regimens may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

USES

Novel DNAs of the present invention can be used to efficiently express a desired protein within a cell. Accordingly, such DNAs can be used in any context in which gene transcription and/or expression is desired.

In one embodiment, the DNA is used in a method of gene therapy to treat a clinical disorder. In another embodiment, the DNA is used in antisense therapy to produce sufficient levels of nuclear and/or cytoplasmic mRNA to inhibit expression of a gene. In another embodiment, the DNA is used to study RNA processing and/or gene regulation *in vitro* or *in vivo*. In another embodiment, the DNA is used to produce therapeutic or diagnostic proteins which can then be administered to patients as exogenous proteins.

Methods for increasing levels of cytoplasmic RNA accumulation and gene expression provided by the present invention can also be used for any and all of the foregoing purposes.

In a preferred embodiment, the invention provides a method if increasing expression of a gene encoding human Factor VIII. Accordingly, the invention also provides an improved method of human Factor VIII gene therapy involving administering to a patient afflicted with a disease characterized by a deficiency in Factor VIII a novel Factor VIII gene in an amount sufficient to treat the disease.

In addition, the present invention provides a novel method for altering the transcription pattern of a DNA. By correcting one or more consensus or near consensus splice sites within the DNA, or by adding one or more introns to the DNA, the natural splicing pattern of the DNA will be modified and, at the same time, expression may be increased. Accordingly, methods of the invention can be used to tailor the transcription of a DNA so that a greater amount of a particular desired RNA species is transcribed and ultimately expressed, relative to other RNA species transcribed from the DNA (i.e., alternatively spliced RNAs).

Methods of the invention can also be used to modify the coding sequence of a given DNA, so that the structure of the protein expressed from the DNA is altered in a beneficial manner. For example, introns can be added to the DNA so that portions of the gene will be removed during transcription and, thus, not be expressed. Preferred gene portions for removal in this manner include those encoding, e.g., antigenic regions of a protein and/or regions not required for activity. Alternatively or additionally, consensus or near consensus splice sites can be corrected within the DNA so that previously

recognizable (i.e., operable) introns and exons are no longer recognized by a cells splicing machinery. This alters the coding sequence of the mRNA ultimately transcribed from the DNA, and can also facilitate its export from the nucleus to the cytoplasm where it can be expressed.

This invention is illustrated further by the following examples which should not be construed as further limiting the subject invention. The contents of all references and published patent applications cited throughout this application are hereby incorporated by reference.

EXAMPLES

EXAMPLE 1 - Construction of a Human Factor VIII Gene Containing an Intron Spanning the β-Domain

A full-length human Factor VIII cDNA containing an intron spanning the section of the cDNA encoding amino acids 745-1638 (Figure 11) was constructed as described below. Amino acid numbering was designated starting with Met-1 of the mature human Factor VIII protein and, thus, does not include the 19 amino acid signal peptide of the protein. The β-domain region of a human Factor VIII protein is made up of 983 amino acids (Vehar et al. (1984) *Nature* 312: 337-342). Thus, the region of the cDNA spliced out during pre-mRNA processing corresponds to about 89% of the β-domain.

To select suitable sites for inserting the 5' splice donor (SD) and 3' splice acceptor (SA) sites, the sequence of the full-length Factor VIII cDNA expression plasmid pCY-6 (SEQ ID NO:4) was scanned for convenient restriction enzyme sites. Restriction sites were selected according to the following criteria: (a) they flanked and were in close proximity to the sites into which the splicing signals were to be introduced, so that any PCR fragment generated to fill in the region between these sites would have as little chance as possible for undesired point mutations introduced by the process of PCR; (b) they would cut the expression plasmid in as few places as possible, preferably only at the site flanking the region of splice site introduction.

The restriction sites chosen according to these criteria for cloning in the splice donor site were: Kpn I (base 2816 of the coding sequence of pCY-6, or base 3822 of the complete nucleotide sequence of pCY-6 provided in SEQ ID NO:4, since the first 1005 bases of this plasmid are non-coding bases), and Tth 1111 (base 3449 of the coding sequence of pCY-6, or base 4455 of the complete nucleotide sequence of pCY-6 shown in SEQ ID NO:4). The restriction sites chosen according to these criteria for cloning in the splice acceptor site were: Bcl I (bases 1407 and 5424 of the coding sequence of pCY-6, or bases 2413 and 6430 of the complete nucleotide sequence of pCY-6 shown in SEQ ID NO:4) and BspE 1 (base 7228 of

the coding sequence of pCY-6, or base 8234 of the complete nucleotide sequence of pCY-6 shown in SEQ ID NO:4).

Generation of Splice Donor Site

A fragment containing the region of Factor VIII cDNA from the Kpn I site to the Tth 111 I site, with the above described splice donor sequence inserted at the appropriate spot, was then generated in the following manner:

A. PCR primers were designed, such that the top strand upstream primer (Fragment A top) would prime at the Kpn I site of full-length Factor VIII cDNA (Figure 12), and the bottom strand downstream primer (Fragment A bottom) would prime at the site of insertion for the 5' splice donor. The bottom strand primer also contained the insertion sequence. These primers were used in a PCR reaction with pCIS-F8 (full-length Factor VIII cDNA expression plasmid) as template to yield "Fragment A," which contains the sequence spanning the region of Factor VIII cDNA from Kpn I to the splice donor insertion site, located at the 3' end of the fragment.

B. In similar fashion, "Fragment B" was generated using primer "Fragment B top," which contains the insertion sequence, and would prime at the insertion site of full-length Factor VIII cDNA, and primer "Fragment B bottom," which would prime at the Tth 111 I site of full-length Factor VIII cDNA. "Fragment B" contains the sequence spanning the region of Factor VIII cDNA from the splice donor insertion site to Tth111 I. The 5' splice donor insertion sequence was located at the 5' end of the fragment.

C. Fragments A and B were run on a horizontal agarose gel, excised, and extracted, in order to purify them away from unincorporated nucleotides and primers.

D. These fragments were then combined in a PCR reaction using as primers "Fragment A top" and "Fragment B bottom." The regions at the 3' end of Fragment A and the 5' end of Fragment B overlapped because they were identical, and the final product of this reaction was a PCR fragment spanning the Factor VIII cDNA from Kpn I to Tth111 I, and containing the engineered splice donor at the insertion site, i.e., near the beginning of the coding region of the β-domain of Factor VIII. This fragment was designated "Fragment AB."

E. Fragment AB (an overlap PCR product) was cloned into the EcoR V site of pBluescript II SK(+) to yield clone pBS-SD (Figure 9), and the sequence of the insertion was then confirmed.

Generation of Splice Acceptor Site

A fragment containing the region of Factor VIII cDNA from the second Bcl I site to the BspE I site, with the above described splice acceptor sequence inserted at the appropriate spot, was generated in the following manner:

A. PCR primers were designed, such that the top strand upstream primer (Primer A) would prime at the second Bcl I site, and the bottom strand downstream primer (Primer B2) would prime at the insertion site for the 3' splice acceptor. The bottom strand primer also contained the restriction sites Mun I and BspE I. These primers were used in a PCR reaction with pCIS-F8 as template to yield "Fragment I," which contains the sequence spanning the region of Factor VIII cDNA from the Bcl I site to the insertion site, with the Mun I and BspE I sites located at the 3' end of the fragment.

B. In a similar fashion, "Fragment III" was generated using "Primer G3" which contains the restriction site BstE II, the splice acceptor recognition sequence (polypyrimidine tract followed by "CAG"), and primes at the insertion site for the splice acceptor; and "Primer H," which would prime the bottom strand at the BspE I site, so that the resulting fragment would contain the restriction site BstE II, the splice acceptor recognition site and sequence spanning the region of Factor VIII cDNA from the insertion site to BspE I.

C. "Fragment II," which contained the branch signals and IVS 14 sequence, was generated by designing four oligos (C2, D, E, and F3), two top and two bottom, which, when combined, would overlap each other by 21 to 22 bases, and when filled in and amplified under PCR conditions, would generate a fragment containing a Mun I site, 130 bases of the aforementioned IVS 14 sequence (including the 2 branch sequences at the 5' end of the 130 bases), and the cloning sites BstE II and BspE I. In addition, two small primers (CX and FX2) were designed that would prime at the very ends of the expected fragment, in order to increase amplification of full-length PCR product. All oligonucleotide primers were combined in a single PCR reaction, and the desired fragment was generated.

D. All three fragments were cloned into the EcoR V site of pBluescript II SK(+), and their sequences were then confirmed.

E. Fragment II was isolated out of pBluescript as a Mun 1 to BspE I fragment, and cloned into the pBluescript-Fragment I clone at the corresponding sites, to yield clone pBS-FI/FII (Figure 9), Fragment III was isolated out of pBluescript as a BstE II to BspE I fragment, and cloned into the corresponding sites of pBS-FI/FII to yield pBS-FI/FII/FIII (Figure 9). This final bluescript clone contained the region spanning Factor VIII cDNA from the second Bcl I site to the BspE I site, and contained the IVS 14 and splice acceptor sequence inserted at the appropriate sites. The pBS-FI/FII/FIII clone was then sequenced.

Cloning Splice Donor and Acceptor Sites into a Factor VIII cDNA Vector (pCY-6)

Fragment AB and Fragment I/II/III were isolated out of pBluescript and cloned into pCY-6 in the following manner:

A. Fragment I/II/III was isolated from pBS-FI/FII/FIII as a Bcl I to BspE I fragment.

B. pCY-601 was digested to completion with BspE I, linearizing the plasmid. This linear DNA was partially digested with Bcl I for 5 minutes, and then immediately run on a gel. The band corresponding to a fragment which had been cut only at the BspE I and the second Bcl I site was isolated and extracted from the agarose gel. This isolated fragment was ligated to Fragment I/II/III and yielded pCY-601/FI/FII/FIII (Figure 9).

C. Fragment AB was isolated from pBS-SD as a Kpn I to Tth111 I fragment, and cloned into the corresponding sites of pCY-601/FI/FII/FIII to yield pLZ-601.

D. Plasmids pCY-6 and pLZ-601 were digested sequentially with enzymes Nco I and Sal I. The small fragment of the pCY-6 digest and the large fragment of the pLZ-601 digest were isolated and ligated together to yield plasmid pLZ-6, a second β-domain intron Factor VIII expression plasmid.

pCY-6 and pCY-601 are expression plasmids for full-length Factor VIII cDNA. The difference between the two is that the former contains an intron in the 5' untranslated region of the Factor VIII transcript, derived from the second IVS of rabbit beta globin gene. The latter lacks this engineered IVS. *In vitro* experiments have shown that pCY-601 yields undetectable levels of Factor VIII, while pCY-6 yields low but detectable Factor VIII levels.

Expression Assays

To test expression of the various Factor VIII cDNA plasmids including those created as described above, plasmids were transfected at a concentration of 2.0-2.5 μg/ml into HuH-7 human carcinoma cells using the calcium phosphate precipitation method described by O'Mahoney et al. (1994) *DNA & Cell Biol.* 13(12): 1227-1232. Expression levels were measured using the KabiCoATest (Kabi Inc., Sweden). This is both a quantitative and a qualitative assay for measuring Factor VIII expression, because it measures enzymatic activity of Factor VIII.

Reverse Transcriptase-PCR Analysis of Cells Transfected With Factor VIII Expression Plasmids

To confirm that the engineered intron spanning the β-domain of the Factor VIII cDNA in plasmid pLZ-6 resulted in proper splicing of the β-domain coding region, reverse transcriptase (RT)-PCR analysis was performed as follows:

HUH7 cells in T-75 flasks were transfected via $CaPO_4$ precipitation with 36 μg of each of the following DNA plasmids:

pCY-2	β-domain deleted human Factor VIII cDNA
pCY-6	Full-length human Factor VIII cDNA
pLZ-6	Full length human Factor VIII cDNA with engineered β-domain intron

75 ng of pCMVhGH was co-transfected as a transfection control. Untransfected cells were grown alongside as a negative control.

Total RNA was isolated from cells 24 hours post-transfection using Gibco BRL Trizol reagent, according to the standard protocol included in product insert.

RT-PCR Experiments were performed as follows: RT-PCR was performed on all RNA preps to characterize RNA. "Minus RT" PCR was performed on all RNA preps as a negative control (without RT, only DNA is amplified). PCR was performed on plasmids used in transfection assays to compare with RT-PCRs of the RNA preps. All RT-PCR was performed with Access RT-PCR system (Promega, Cat. #A1250). In each 50 µl reaction, 1.0 µg total RNA was used as template. Primer pairs were designed according to Factor VIII sequences as follows: the 5' primer anneals to the top strand of Factor VIII, about 250 base pairs upstream of the β-domain junction; while the 3' primer anneals to the bottom strand of Factor VIII, about 250 base pairs downstream of the β-domain junction.

The nucleotide sequences of the primers used to characterize (i.e., confirm) the β-domain intron splicing were as follows:

5' primer	TS 2921-2940: (20 mer)	$^{5'}$TGG TCT ATG AAG ACA CAC TC$^{3'}$
3' primer	BS 6261-6280: (20 mer)	$^{5'}$TGA GCC CTG TTT CTT AGA AC$^{3'}$

RT-PCR files were set up according to manufacturer's recommendation:

48°C, 45 minutes; x1 cycle
94°C, 2 minutes; x1 cycle
94°C, 30 sec; x 40 cycles
60°C, 1 min; x 40 cycles
68°C, 2 min; x 40 cycles
68°C, 7 min; x 1 cycle
4°C, soak overnight

The data obtained from the RT-PCR assays demonstrated that engineered β-domain intron was spliced as predicted. The RT-PCR product (~500 bp) generated from pLZ-6 (containing the β-domain intron) was similar to that obtained from pCY-2 (containing β-domain deleted Factor VIII cDNA). The RT-PCR product observed for pCY-6 (containing the full length Factor VIII cDNA) yielded a much larger band (~3.3 kb).

In the control groups, it was confirmed that DNA from the Huh-7 cells transfected with various Factor VIII constructs were consistent with regular PCR results of the corresponding plasmids. Background bands from untransfected Huh-7 cells were presumably contributed by cross-over during sample handling. This can be further investigated by using polyA$^+$ RNA as template, as well as by setting up RT-PCR with different primer sets.

EXAMPLE 2 - Correction of Consensus and near Consensus Splice Sites Within a Human Factor VIII Gene

Plasmid pCY-2, containing the coding region of the β-domain deleted human Factor VIII cDNA (nucleotides 1006-5379 of SEQ ID NO:2), was analyzed using the MacVector™ program for consensus and near consensus (a) splice donor sites, (b) splice acceptor sites and (c) branch sequences. Near consensus 5' splice donor sites were selected using the following criteria: sites were required to contain at least 5 out of the 9 splice donor consensus bases (i.e., (C/A)AGGT(A/G)AGT), including the invariant GT, provided that if only 5 out of 9 bases were present, these 5 bases were located consecutively in a row. Near consensus 3' splice acceptor sites were selected using the following criteria: sites were required to contain at least 3 out of the following 14 splice acceptor consensus bases (Y=10)CAGG (wherein Y is a pyrimidine within the pyrimidine track), including the invariant AG. Only branch sequences which were 100% consensus were searched for.

Using these criteria, 23 near consensus 5' splice donor sequences, 22 near consensus 3' splice acceptor sequences, and 18 consensus branch sequences were identified. No consensus 5' splice donor or 3' splice acceptor sequences were identified. To correct these near consensus splice donor and acceptor sequences, and consensus branch sequences, it was first determined whether the invariant GT, AG, or A bases within the site could be substituted without changing the coding sequence of the site. If they could be, then these conservative (silent) substitutions were made, thereby rendering the site non-consensus (since the invariant bases are required for recognition as a splice site).

If the invariant bases within selected consensus and near consensus sites could not be substituted without changing the coding sequence of the site (i.e., if no degeneracy existed for the amino acid sequence coded for), then the maximum number of silent point mutations were made to render the site as far from consensus as possible. All bases which contributed to homology of the consensus or near consensus site with the corresponding consensus sequence, and which were able to be conservatively substituted (with non-consensus bases), were mutated.

Using these guidelines, 99 silent point mutations were selected, as shown in Figure 4A-4C. The positions of each of these silent point mutations is shown in Figure 3.

To prepare a new pCY-2 human β-domain deleted Factor VIII cDNA coding sequence which contains the above-described corrections, the following procedure can be used:

Overlapping 60-mer oligonucleotides can be synthesized based on the coding sequence of pCY2. Each of the 185 oligonucleotide contains the desired corrections. These oligonucleotides are then assembled in five segments (shown in Figure 9) using the method of Stemmer et al. (1995) *Gene* 164: 49-53. Prior to assembly, each segment can be

sequenced and tested in *in vitro* transfection assays (nuclear and cytoplasmic RNA analysis) in pCY2. A schematic illustration of this process is shown in Figures 8. The plasmid containing the new corrected coding sequence is desginated "pDJC."

To test expression levels of pDJC, the plasmid can be transfected at a concentration of 2.0-2.5 μg/ml into HuH-7 human carcinoma cells using any suitable transfection technique, such as the calcium phosphate precipitation method described by O'Mahoney et al. (1994) *DNA & Cell Biol.* 13(12): 1227-1232. Factor VIII expression can then be measured using the KabiCoATest (Kabi Inc., Sweden). This is both a quantitative and a qualitative assay for measuring Factor VIII expression, because it measures enzymatic activity of Factor VIII. Alternatively, plasmids such as pDJC can be tested for *in vivo* expression using the procedure described below in Example 4.

EXAMPLE 3 - Optimized Expression Vectors

Optimized expression vectors for liver-specific and endothelium-specific human Factor VIII expression were prepared and tested as follows:

The β-domain deleted human Factor VIII cDNA was obtained through Bayer Corporation in plasmid p25D, having a coding sequence corresponding to nucleotides 1006-5379 of SEQ ID NO:2. The human thyroid binding globulin promoter (TBG) (bases -382 to +3) was obtained by PCR from human liver genomic DNA (Hayashi et al. (1993) *Mol. Endo.* 7:1049). The human endothelin-1 (ET-1) gene promoter (Lee et al. (1990) *J. Biol. Chem.* 265(18) was synthesized by amplification of overlapping oligos in a PCR reaction.

After sequence confirmation, the TBG and ET-1 promoters were cloned into two separate vectors upstream of an optimized leader sequence (SEQ ID NO:11), using standard cloning techniques. The leader sequence was designed in a similar manner to that reported by Kozak et al. (1994) *J. Mol. Biol.* 235:95) and synthesized (Retrogen Inc., San Diego, CA) as 71 base pair top and bottom strand oligos, annealed and cloned upstream of the Factor VIII ATG. The 126 base pair intron-1 of the rabbit β-globin gene, containing the nucleotide sequence modifications shown in Figure 23 (SEQ ID NO:7), was also synthesized and inserted into the leader sequence following base 42 of the 71 nucleotide sequence.

In the construct containing the TBG promoter, top and bottom strands of the human alpha-1 microglobulin/bikunin enhancer (ABP), sequences -2804 through -2704 (Rouet et al. (1992) *J. Biol. Chem.* 267:20765), were synthesized, annealed and cloned upstream of the promoter. Cloning sites flanking the enhancer were designed to facilitate easy multimerization. In the construct containing the ES-1 promoter, top and bottom strands of the human c-fos SRE enhancer (Treisman et al. (1986) *Cell* 46) were synthesized, annealed and cloned upstream of the promoter.

The post-transcriptional regulatory element (PRE) from hepatitis B virus, was isolated from plasmid Adw-HTD as a 587 base-pair Stu I-Stu I fragment. It was cloned into the 3'

UTR of the Factor VIII construct (at the Hpa I site) containing the TBG promoter and ABP enhancers, upstream of the polyadenylation sequence. A two copy PRE element was isolated as a Spe I-Spe I fragment from an early vector where two copies had ligated together. This fragment was converted to a blunt end fragment by the Klenow fragment of E-coli DNA polymerase I and also cloned into the Factor VIII construct at the same Hpa I site.

Thus, the following constructs were produced using the foregoing materials and methods:

Plasmid pCY-2 having a 5' untranslated region containing the TBG promoter, two copies of the ABP enhancer; and the modified rabbit β-globin IVS, all upstream of the human β-domain deleted Factor VIII gene.

Plasmid pCY2-SE5 which was identical to pCY-2, except that the TBG promoter was replaced by the ET-1 gene promoter, and the ABP enhancers (both copies) were replaced by one copy of the SRE enhancer.

Plasmid pCY-201 which was identical to pCY-2, except that it lacked the 5' intron.

Plasmid pCY-401 and pCY-402 which were identical to pCY-201, except that they contained one and two copies of the HBV PRE, respectively.

Expression levels for each of the foregoing gene constructs was compared in human hepatoma cells (HUH-7) maintained in DMEM (Dulbecco's modified Eagle medium (GIBCO BRL), supplemented with 10% heat inactivated fetal calf serum (10% FCS), penicillin (50 IU/ml), and streptomycin (50 μg/ml) in a humidified atmosphere of 5% CO_2 at 37°C. For experiments involving quantitation of human factor VIII protein, media was supplemented with an additional 10% FCS. DNA transfection was performed by a calcium phosphate coprecipitation method.

Other human Factor VIII gene constructs (shown below in Table I) tested for expression, prepared as described above, included constructs which were identical to pCY-2, except that they contained (a) the TBG promoter with no enhancer or 5' intron, (b) the TBG promoter with a 5' modified rabbit β-globin intron (present within the leader sequence), but no enhancer, (c) the TBG promoter with one copy of the ABP enhancer and a 5' modified rabbit β -globin intron (present within the leader sequence), and (d) the TBG promoter with two copies of the ABP enhancer and a 5' modified rabbit β-globin intron (present within the leader sequence).

Active Factor VIII protein was measured from tissue culture supernatants by COAtest VIII:c/4 kit assay specific for active Factor VIII protein. Transfection efficiencies were normalized to expression of cotransfected human growth hormone (hGH).

As shown below in Table I, liver-specific human Factor VIII expression is significantly increased by the combined use of the TBG promoter and a 5' intron within the 5' UTR of the gene construct. Expression is further increased (over 30 fold) by adding a copy of the ABP enhancer in the same construct. Expression is still further increased (over 60 fold) by

using two copies of the ABP enhancer in the same construct. In addition, as shown in Figure 18, expression is also significantly increased by adding one or more PRE sequences into the 3' UTR of the gene construct, although, in this experiment, not as much as by adding a 5' intron within the 5' UTR.

TABLE I

5' Region Tested	Fold Increase in Factor VIII Expression *In Vitro*
TBG Promoter	1
TBG Promoter, 5' Intron	3.5
ABP Enhancer (1 copy), TBG Promoter, 5' Intron	30.1
ABP Enhancer (2 copies), TBG Promoter, 5' Intron (pCY-2)	63.2

Expression of pCY2-SE5 was also tested and compared with pCY-2 in (a) bovine aortic endothelial cells and (b) HUH-7 cells. Transfections and Assays were performed as described above. Significantly more biologically active human Factor VIII was secreted from cells transfected with pCY2-SE5 than with pCY-2 (625 pg/ml vs. 280 pg/ml). While liver-specific pCY-2 expressed more than 10 ng/ml of human Factor VIII from HUH-7 cells, no human Factor VIII could be detected from pCY2-SE5 transfected HUH-7 cells.

Constructs were also tested *in vivo*. Specifically, pCY-2 and pCY2-SE5 were tested in mouse models by injecting mice (tail vein) with 10 μg of DNA in one 1.0 ml of solution (0.3 M NaCl, pH 9). Plasmids pCY-6, pLZ-6 and pLZ-6A (described in Example 1) were tested in the same experiment. Levels of human Factor VIII were measured in mouse serum. The results are shown in Figure 19. Plasmid pCY-2, containing the TBG promoter, 2 copies of the ABP enhancer, and an optimized 5' intron, had the highest expression, followed by pLZ-6A, pLZ-6, pCY2-SE5 and pCY-6.

Plasmid pCY-2 was also tested *in vivo* in mice, along with plasmid p25D which contained the same coding sequence (for human β-domain deleted Factor VIII) without an optimized 5' UTR. Specifically, instead of 2 copies of the ABP enhancer, one copy of the TBG promoter and a leader sequence containing an optimized (i.e., modified to contain consensus splice donor and acceptor sites and a consensus branch and pyrimidine track sequence) 5' rabbit β-globin intron (as contained in the 5' UTR of pCY-2), p25D contained within its 5' UTR one copy of the CMV enhancer, one copy of the CMV promoter, and a

leader sequence containing an unmodified short (130 bp) chimeric human IgE intron (containing uncorrected near consensus splice donor and acceptor sites). Plasmids were injected into mice (tail vein) in the form of asialoorosomucoid/polylysine/DNA complexes formed as described below in Example 4. Mice were injected with 10 μg of DNA (complexed) in 1.0 of solution (0.3 M NaCl, pH 9).

The results are shown in Figure 25 and demonstrate that optimization of gene constructs by modification of 5' UTRs to contain novel combinations of strong tissue-specific promoters and enhancers, and optimized introns (e.g. modified to contain consensus splice donor and acceptor sites and a consensus branch and pyrimidine track sequence) significantly increases both levels and duration of gene expression. Notably, expression of p25D shut off after only 8 days, whereas expression of pCY-2 was maintained at nearly 100% of initial levels (well in the human therapeutic range of 10 ng/ml or more) for over 10 days. In the same experiment, expression was maintained well in the therapeutic range for greater than 30 days.

Overall, the results of the foregoing examples demonstrate that gene expression can be significantly increased and prolonged *in vivo* by optimizing untranslated regulatory regions and/or coding sequences in accordance with the teachings of the present invention.

EXAMPLE 4 - Targeted Delivery of Novel Genes to Cells

Novel genes of the invention, such as novel Factor VIII genes contained in appropriate expression vectors, can be selectively delivered to target cells either *in vitro* or *in vivo* as follows:

Formation of Targeted Molecular Complexes

I. *Reagents*

Protamine, poly-L-lysine (4kD, 10kD, 26kD; mean MW) and ethidium bromide can be purchased from Sigma Chemical Co., St. Louis, MO. 1-[3-(dimethylamino)-propyl]-3-ethylcarbodiimide (EDC) can be purchased from Aldrich Chemical Co, Milwaukee, WI. Synthetic polylysines can be purchased from Research Genetics (Huntsville, AL) or Dr. Schwabe (Protein Chemistry Facility at the Medical University of South Carolina). Orosomucoid (OR) can be purchased from Alpha Therapeutics, Los Angeles, CA. Asialoorosomucoid (AsOR) can be prepared from orosomucoid (15 mg/ml) by hydrolysis with 0.1 N sulfuric acid at 76°C for one hour. AsOR can then be purified from the reaction mixture by neutralization with 1.0 N NaOH to pH 5.5 and exhaustive dialysis against water at room temperature. AsOR concentration can be determined using an extinction coefficient of 0.92 ml mg^{-1}, cm^{-1} at 280 nm. The thiobarbituric acid assay of Warren (1959) *J. Biol. Chem.* 234:1971-1975 or of Uchida (1977) *J. Biochem.* 82:1425-1433 can be used to verify desialylation of the OR. AsOR prepared by the above method is typically 98% desialyated.

II. *Formation of Carrier Molecules*

Carrier molecules capable of electrostatically binding to DNA can be prepared as follows: AsOR-poly-L-lysine conjugate (AP26K) can be formed by carbodiimide coupling similar to that reported by McKee (1994) *Bioconj. Chem.* 5:306-311. AsOR, 26kD poly-L-lysine and EDC in a 1:1:0.5 mass ratio can be reacted as follows. EDC (dry) is added directly to a stirring aqueous AsOR solution. Polylysine (26 kD) is then added, the reaction mixture adjusted to pH 5.5-6.0, and stirred for two hours at ambient temperature. The reaction can be quenched by addition of Na_3PO_4 (200 mM, pH 11) to a final concentration of 10 mM. The AP26K conjugate can be first purified on a Fast Flow Q Sepharose anion exchange chromatography column (Pharmacia) eluted with 50 mM Tris, pH 7.5; and then dialyzed against water.

III. *Calculation of Charge Ratios* (+/-)

Charge ratios of purified carrier molecules can be determined as follows: Protein-polylysine conjugates (e.g., AsOR-PL or OR-PL) are exhaustively dialyzed against ultra-pure water. An aliquot of the dialyzed conjugate solution is lyophilized, weighed and dissolved in ultra-pure water at a specific concentration (w/v). Since polylysine has minimal absorbance at 280 nm, the AsOR component of AsOR-polylysine (w/v) is calculated using the extinction coefficient at 280 nm. The composition of the conjugate is estimated by comparison of the concentration of the conjugate (w/v) with the concentration of AsOR (w/v) as determined by UV absorbance. The difference between the two determinations can be attributed to the polylysine component of the conjugate. The composition of OR-polylysine can be calculated in the same manner. The ratio of conjugate to DNA (w/w) necessary for specific charge ratios then can be calculated using the determined conjugate composition. Charge ratios for molecular complexes made with, e.g., polylysine or protamine, can be calculated from the amino acid composition.

IV. *Complexation With DNA*

To form targeted DNA complexes, DNA (e.g., plasmid DNA) is preferably prepared in glycine (e.g., 0.44 M, pH 7), and is then rapidly added to an equal volume of carrier molecule, also in glycine (e.g., 0.44 M, pH 7), so that the final solution is isotonic.

V. *Fluorescence Quenching Assay*

Binding efficiencies of DNA to various polycationic carrier molecules can be examined using an ethidium bromide-based quenching assay. Solutions can be prepared containing 2.5 μg/ml EtBr and 10 μg/ml DNA (1:5 EtBr:DNA phosphates molar ratio) in a total volume of 1.0 ml. The polycation is added incrementally with fluorescence readings

taken at each point using a fluorometer (e.g., a Sequoia-Turner 450), with excitation and emission wavelengths at 540 nm and 585 nm, respectively. Fluorescence readings are preferably adjusted to compensate for the change in volume due to the addition of polycation, if the polycation did not exceed 3% of the original volume. Results can be reported as the percentage of fluorescence relative to that of uncomplexed plasmid DNA (no polycation).

Cell Delivery *In Vivo* or *In Vitro*

DNA complexes prepared as described above can be administered in solution to subjects via injection. By way of example, a 0.1-1.0 ml dose of complex in solution can be injected intravenously via the tail vein into adult (e.g., 18-20 gm) BALB/C mice, at a dose ranging from <1.0-10.0 μg of DNA complex per mouse.

Alternatively, DNA complexes can be incubated with cells (e.g., HuH cells) in culture using any suitable transfection protocol known in the art for targeted uptake. Target cells for transfection must contain on their surface a component capable of binding to the cell-binding component of the DNA complex.

EQUIVALENTS

Although the invention has been described with reference to its preferred embodiments, other embodiments can achieve the same results. Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific embodiments described herein. Such equivalents are considered to be within the scope of this invention and are encompassed by the following claims.

INCORPORATION BY REFERENCE

The contents of all references and patents cited herein are hereby incorporated by reference in their entirety.